



Vyhodnocení dotazníkového šetření pomocí korespondenční analýzy

Julie POLÁČKOVÁ, Andrea JINDROVÁ, Czech University of Life Sciences Pragueⁱ

Abstract

The article is focused on the evaluation of a questionnaire survey. The aim was to determine whether the creation of interactive manuals, which monitor working in statistical software, helps to make education processes effective. For a clear evaluation and presentation of the pilot survey a method of correspondence analysis was chosen. The correspondence analysis is a multivariate graphical technique for representing information in a two-way contingency table. Analysis outputs indicate that these manuals are an effective way of teaching how to work in statistical software. This conclusion was confirmed primarily by less skilful students.

Keywords

Categorical data analysis, correspondence analysis, innovative teaching, interactive presentation, questionnaire survey analysis.

JEL Classification: C49, C83

ⁱ Department of Statistics, Faculty of Economics and Management, Czech University of Life Sciences, Kamýcká 129, 165 21 Prague 6 – Suchbát, Czech Republic.

polackova@pef.czu.cz (corresponding author)

This research was supported by the Science Foundation (IGA) of the Czech University of Life Sciences under the project no. 200911170007.

1. Úvod

Vytvoření kvalitní studijní opory představuje metodický aparát, který výrazně zefektivní výuku statistických předmětů využívajících statistické programy. Nejen jazyková bariéra, ale i technická náročnost některých programů s sebou přináší potřebu zhotovení názorných návodů. V rámci zefektivnění výuky statistických předmětů, především výuky práce ve statistických systémech, jsou vytvářeny názorné návody (instruktační prezentace), které demonstrují postupy práce ve statistických programech. Záměrem autorů těchto prezentací je usnadnit a přiblížit studentům ovládání těchto systémů a orientaci v nich. Pomocí realizovaného pilotního dotazníkového šetření je zkoumáno, zda jsou tyto prezentace studenty vnímány jako kvalitní studijní podpora, má-li jejich využívání vliv na zefektivnění domácí přípravy a zda studenti

dokáží na základě těchto manuálů pochopit postupy tvorby statistických analýz.

Cílem příspěvku je analyzovat získané informace z dotazníkového šetření pomocí korespondenční analýzy. První část příspěvku je zaměřena na zkoumání existence spojitostí mezi vztahem studentů ke studiu a jejich názorem na užitečnost realizovaných prezentací. V druhé části je zkoumán vztah mezi zkušenostmi studentů s informačními technologiemi a jejich názorem na zefektivnění domácí přípravy. Je možné předpokládat, že prospěšnost realizovaných návodů ocení především uživatelé s nižší počítačovou gramotností, kteří vykazují větší problémy s orientací v počítačových systémech.

2. Použitá metodika

K analýze dotazníkového šetření byly využity nástroje korespondenční analýzy dat. Tato analýza představuje

populární grafickou techniku využívanou k analýze vztahů mezi kategoriemi jedné či více proměnných v kontingenčních tabulkách. Pomocí nástrojů korespondenční analýzy je možné popsat asociace nominálních či ordinálních proměnných a získat grafické znázornění souvislosti ve vícerozměrném prostoru (více Ramos a Carvalho, 2010). Beh (2010) spatřuje největší výhodu této metody právě v její schopnosti graficky znázornit propojenost jednotlivých kategorií. Rencher (2002) zdůrazňuje, že základem vytváření subjektivní (korespondenční) mapy jsou tzv. latentní veličiny. Polohy bodů v subjektivní mapě přímo vyjadřují asociaci; vzdálenosti mezi body (neboli vzdálenost řádkových a sloupcových profilů) je možné přenést do dvojrozměrné euklidovské roviny, ve které body odpovídají jednotlivým kategoriím.

Hebák a kol. (2007) dodává, že korespondenční analýza zobrazuje korespondence kategorií jednotlivých proměnných a poskytuje společný obraz řádkových i sloupcových kategorií ve stejných dimenzích. Na rozdíl od většiny ostatních vícerozměrných metod umožňuje korespondenční analýza zpracování kategorizovaných nemetrických dat i nelineárních vztahů. Představuje obdobu faktorové analýzy, místo faktorů je však sledován vliv jednotlivých kategorií, jejich vzájemná podobnost či asociace s kategoriemi ostatních proměnných (Rencher, 2002).

Cílem korespondenční analýzy je dle Hebáka a kol. (2007) ... *redukce mnohorozměrného prostoru vektorů řádkových a sloupcových profilů při maximálním zachování informace obsažené v původních datech*. V subjektivním mapování bývá nejčastěji využíváno dvojrozměrného (roviny) či maximálně trojrozměrného zobrazení vzdáleností v euklidovském prostoru. Častěji než euklidovská vzdálenost se využívá Pearsonova statistika chí-kvadrát. Blízké řádkové body indikují řádky, které mají podobné profily v celém řádku, blízké sloupcové body indikují sloupce s podobnými profily směrem dolů přes všechny řádky. A řádkové body, které jsou v těsné blízkosti sloupcových bodů, představují kombinace, které se objeví častěji, než by se očekávalo u nezávislého modelu, ve kterém řádkové kategorie nejsou vztaženy ke sloupcovým (Meloun a kol., 2005).

Rozptýlenost bodů je možné posuzovat dle ukazatele inercie, který odpovídá váženému průměru chí-kvadrát vzdáleností řádkových (respektive sloupcových) profilů od svého průměru (Meloun a kol., 2005). Singulární hodnota a inercie odpovídá vlastnímu číslu v analýze hlavních komponent; představuje míru variability mezi profily vysvětlenou danou dimenzí řešení nebo danou kategorií. Podle toho určíme potřebný počet dimenzí. Odlišnost profilů, měřená pomocí míry založené na chí-kvadrát statistice, je to, co se projeví v grafu jako vzdálenost mezi položkami

stejně proměnné. Vzdálenost mezi položkami různých proměnných jsou obrazy standardizovaných reziduí na průsečíku položek.

Jako hlavní předpoklad pro použití korespondenční analýzy uvádí autoři Meloun a kol. (2005) kromě porovnatelnosti objektů také úplnost datové matice. Řešení vychází z matice standardizovaných reziduí, kterou je možné vytvořit na základě některé z normalizačních metod. Výběr normalizační metody závisí na preferencích výzkumníka. Při preferencích vztahů mezi řádkovými kategoriemi je využívána analýza řádkových profilů, při upřednostnění sloupcových kategorií vycházíme z analýzy sloupcových profilů. Kombinací těchto dvou analýz je metoda symetrická, která umožňuje vzájemné srovnání řádkové a sloupcové kategorie. Tato metoda je preferována, pokud je cílem vytvořit bodový graf sloupcových a řádkových profilů (symetrické mapy) (Řezanková, 2007).

Pilotního dotazníkového šetření se zúčastnili studenti 4. ročníku oboru Informatika (INFO) a studenti 2. ročníku oboru Podnikání a administrativa (PaA), kteří využívají statistický software v rámci výuky statistických předmětů. Distribuce a sběr dotazníků probíhaly dvojí formou. U oboru INFO byly dotazníky vyplňovány elektronicky v příslušném kurzu vytvořeném v systému Moodle. U oboru PaA byly dotazníky předány studentům k vyplnění přímo v jednotlivých cvičeních v papírové podobě.

Ke zpracování dotazníkového šetření byl využit statistický software SPSS Statistics 17.0 for Windows, který představuje moderní systém pro popis a vizualizaci datových vztahů. Výhodnou daného softwaru při zpracování dat z dotazníkových šetření je existence tzv. codebooku, který obsahuje informace o datové matici (názvy a popisy proměnných, popisy hodnot, chybějící hodnoty), čímž uživateli usnadní následnou realizaci statistických analýz.

3. Vlastní zpracování a diskuze

Pomocí realizovaného pilotního dotazníkového šetření byla zjišťována efektivita vytvořených interaktivních prezentací. Cílem bylo získání poznatků o tom, zda vytvoření komplexního studijního materiálu demonstrovajícího práci ve statistickém softwaru je pro studenty přínosné a přispěje k jejich lepší a rychlejší orientaci v daném systému.

3.1 Analýza vztahu studentů ke studiu a jejich názoru na instruktážní prezentace

Nejprve byla analyzována závislost vztahu studentů ke studiu a jejich názoru na tvrzení, že jim interaktivní prezentace (IP) umožnily překonat potíže s orientací ve statistickém systému.

Tabulka 1 zachycuje vstupní absolutní četnosti. Počet kategorií u obou proměnných je roven třem, z čehož je možné odvodit maximální počet dimenzí použitých v proceduře. Maximální počet dimenzí představuje nejmenší počet kategorií u jednotlivých proměnných snížený o jednotku. Počet dimenzí bude tedy roven dvěma.

Ze souhrnné tabulky korespondenční analýzy vyplynulo, že první dimenze zachycuje 89 % celkové informace – variability (inerce), druhá dimenze vysvětluje zbývajících 11 % informace.

Ukazatel zátěží (Mass), uvedený v tabulce 2, označuje řádkovou zátěž, která představuje procentuální podíl informace z celé tabulky v dané kategorii. Tuto zátěž získáme jako poměr řádkové či sloupcové okrajové četnosti (n_{i+} , n_{+j}) a celkové četnosti tabulky (n). Nejvyšší podíl řádkové proměnné nalezneme u třetí kategorie, která zastupuje výrok *určitě souhlasím*. Kategorie *připravuji se pouze v rámci zadaných projektů* představuje nejvyšší podíl sloupcové proměnné. Následující dva sloupce (Score in Dimension) zachycují skóre v prvních dvou rozměrech.

Ukazatel inerce (Inertia) představuje podíl celkové informace na profilu čili na dotyčném bodu. Tato charakteristika je nezávislá na počtu dimenzí.

Z tabulky 2 je patrné, že první hlavní komponenta v řádkovém profilu dělí studenty na ty, kteří s tvrzením, že jim IP umožnily překonat potíže s orientací ve statistickém systému, souhlasí, a ty, co jsou neutrální, tedy souhlasí pouze částečně. Druhá hlavní komponenta rozděluje studenty, kterým interaktivní prezentace umožnily překonat potíže, dle míry souhlasu.

První hlavní komponenta ve sloupcovém profilu (viz tabulka 3) odděluje studenty, kteří se na studium připravují, od těch, kteří preferují tzv. *cestu nejmenšího odporu*. Druhá hlavní komponenta dělí studenty dle míry zaujetí studiem.

K zobrazení korespondenční mapy, která zachycuje grafické znázornění sloupcových i řádkových kategorií, byl zvolen symetrický model. Na rozdíl od asymetrických map, ve kterých docházelo k přehuštění prostoru a celkové nečitelnosti modelu, jsou v symetrických mapách řádkové a sloupcové body rozptýleny rovnoměrně.

Tabulka 1 Korespondenční tabulka absolutních četností

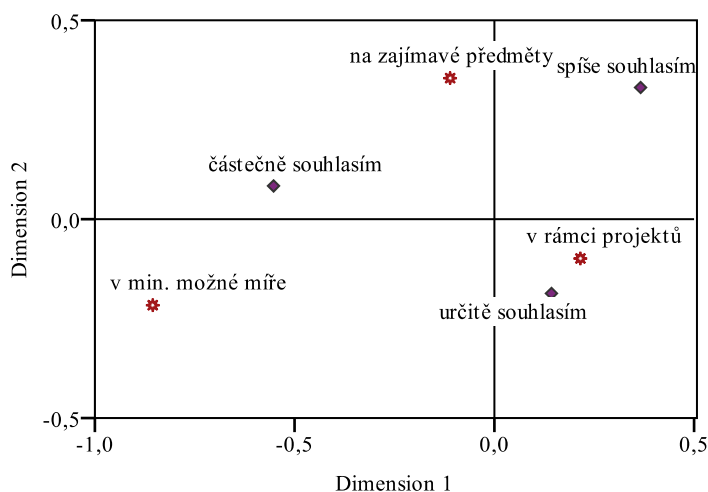
IP mi umožnily překonat potíže s orientací ve statistickém systému	Jaký je Váš vztah ke studiu? Připravuji se...			
	na zajímavé předměty	v rámci projektů	v minimální možné míře	Active Margin
částečně souhlasím	9	18	6	33
spíše souhlasím	7	17	2	26
určitě souhlasím	14	40	7	61
Active Margin	30	75	15	120

Tabulka 2 Tabulka řádkových zátěží

IP mi umožnily překonat potíže s orientací ve statistickém systému		Score in Dimension			Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
		Mass	1		2	Inertia	1	2	1
částečně souhlasím	0,275	−0,553	0,084	0,010	0,681	0,044	0,992	0,008	1,000
spíše souhlasím	0,217	0,366	0,331	0,005	0,235	0,548	0,777	0,223	1,000
určitě souhlasím	0,508	0,143	−0,186	0,002	0,084	0,407	0,626	0,374	1,000
Active Total	1,000			0,017	1,000	1,000			

Tabulka 3 Tabulka sloupcových zátěží

Jaký je Váš vztah ke studiu? Připravuji se...		Score in Dimension			Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
		Mass	1		2	Inertia	1	2	1
na zajímavé předměty	0,250	−0,111	0,355	0,002	0,025	0,725	0,217	0,783	1,000
v rámci projektů	0,625	0,215	−0,099	0,004	0,235	0,140	0,931	0,069	1,000
v minimální možné míře	0,125	−0,855	−0,216	0,012	0,740	0,135	0,978	0,022	1,000
Active Total	1.000			0.017	1.000	1.000			



- ◆ IP mi umožnily překonat potíže s orientací ve statistickém systému.
- * Jaký je Váš vztah ke studiu? Připravuji se...

Obrázek 1 Symetrická korespondenční mapa sloupcových a řádkových profilů

Výstupy korespondenční analýzy zahrnují názorné grafy, které ilustrují vztahy mezi jednotlivými kategoriemi a proměnnými. Následující bodový graf zachycuje řádkové a sloupcové skóre pro dvojdimenzionální řešení. Využití symetrické normalizace zjednodušuje prozkoumání vztahů mezi jednotlivými kategoriemi daných proměnných.

Obrázek 1 zachycuje vzájemné vztahy zkoumaných kategorií. Z grafu jsou patrné tři oddělené skupiny kategorií. Studenti, kteří se připravují především na zajímavé předměty, spíše souhlasí s tvrzením, že jim IP umožnily překonat potíže s orientací ve statistickém systému. Studenti, kteří se připravují pouze v rámci zadaných projektů, s daným tvrzením souhlasí určitě, jelikož jim IP z pravidla usnadní vytvoření projektu. Naopak studenti, kteří se na výuku připravují v minimální možné míře, s tvrzením souhlasí pouze částečně, jelikož raději zvolí rychlejší a méně podrobné formy nápovědy.

3.2 Analýza vztahu mezi IT zkušenostmi a názorem na zefektivnění domácí přípravy

Dále byl zkoumáno, zda existuje vztah mezi zkušenostmi studentů s informačními technologiemi a názorem na zefektivnění domácí přípravy. Lze

předpokládat, že výhody realizovaných prezentací ocení především studenti s nižší zkušeností s IT. Absolutní četnosti jednotlivých kategorií jsou uvedeny v tabulce 4.

Stejně jako v první analýze, i zde byly využity pouze dvě dimenze. Ze souhrnné tabulky korespondenční analýzy vyplynulo, že první dimenze zachycuje 88 % celkové informace a druhá dimenze vysvětluje zbývajících 12 % informace.

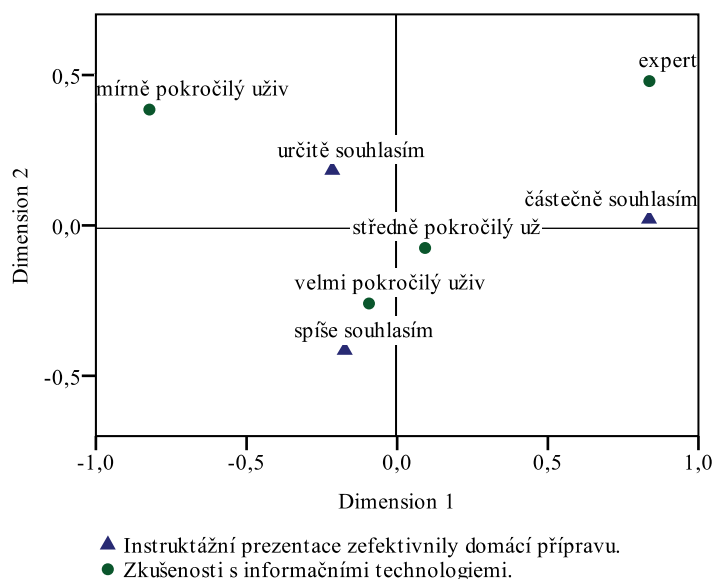
První hlavní komponenta v řádkovém profilu dělí studenty na ty, kteří souhlasí s tvrzením, že IP zefektivnily domácí přípravu, zbývajících studenti jsou neutrální, tedy souhlasí pouze částečně. Druhá hlavní komponenta rozděluje studenty dle míry souhlasu s výrokem, že využití interaktivních prezentací zefektivnilo jejich domácí přípravu (viz tabulka 5).

Z tabulky 6 je patrné, že první hlavní komponenta ve sloupcovém profilu dělí studenty dle zkušenosti s informačními technologiemi na uživatele a experty (administrátory apod.). Druhá hlavní komponenta pak rozděluje studenty dle zkušenosti na mírně, středně a velmi pokročilé uživatele.

Vzájemné vztahy zkoumaných kategorií zachycuje obrázek 2. Z grafu již nejsou patrné tak výrazně

Tabulka 4 Korespondenční tabulka absolutních četností

IP zefektivnily domácí přípravu	zkušenosti s IT				
	mírně pokročilý uživatel	středně pokročilý uživatel	velmi pokročilý uživatel	expert	Active Margin
částečně souhlasím	1	13	5	4	23
spíše souhlasím	4	16	8	2	30
určitě souhlasím	11	33	15	6	65
Active Margin	16	62	28	12	118

**Obrázek 2** Symetrická korespondenční mapa sloupkových a řádkových profilů

oddělené skupiny kategorií, jak tomu bylo v předchozím případě. Obecně lze ale konstatovat, že čím je uživatel pokročilejší, tím méně zastává názor, že IP ztrátlivnily jeho domácí přípravu. S daným tvrzením určitě souhlasí převážně méně pokročilí uživatelé. Velmi pokročilí uživatelé s tímto názorem souhlasí spíše a experti souhlasí pouze částečně.

Získané výsledky podporují záměry autorů instruktažních prezentací, kteří se zaměřili především na méně zkušené uživatele, kteří mají větší problémy s orientací ve složitých programových systémech. Pokročilí uživatelé či programátoři již pravděpodobně

ne mají s orientací v nových programech větší potíže, nebo si v případě potřeby naleznou potřebné informace ve zdrojích dostupných na internetu či v nápovědách.

4. Závěr

Z uvedených analýz vyplývá, že studenti, kteří se věnují domácí přípravě intenzivněji, více oceňují existenci instruktažních prezentací, které jim pomáhají překonat problémy s orientací ve statistickém systému. Studenti s jejich pomocí snáze porozumí práci ve

Tabulka 5 Tabulka řádkových zátěží

IP zefektivnily domácí přípravu		Score in Dimension			Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
	Mass	1	2	Inertia	1	2	1	2	Total
částečně souhlasím	0,195	0,836	0,021	0,023	0,804	0,001	1,000	0,000	1,000
spíše souhlasím	0,254	−0,174	−0,415	0,004	0,045	0,700	0,323	0,677	1,000
určitě souhlasím	0,551	−0,216	0,184	0,006	0,151	0,298	0,789	0,211	1,000
Active Total	1,000			0,033	1,000	1,000			

Tabulka 6 Tabulka sloupkových zátěží

zkušenosti s IT		Score in Dimension			Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
	Mass	1	2	Inertia	1	2	1	2	Total
mírně pokročilý uživatel	0,136	−0,822	0,385	0,017	0,541	0,322	0,925	0,075	1,000
středně pokročilý uživatel	0,525	0,093	−0,075	0,001	0,027	0,047	0,805	0,195	1,000
velmi pokročilý uživatel	0,237	−0,094	−0,259	0,001	0,012	0,256	0,262	0,738	1,000
expert	0,102	0,837	0,480	0,014	0,420	0,375	0,892	0,108	1,000
Active Total	1,000			0,033	1,000	1,000			

statistických systémech. Využití těchto prezentací v rámci samostudia vede k celkovému zefektivnění domácí přípravy. Tento fakt je přínosný nejen z pohledu studentů, ale i z hlediska pedagogů, kteří mají na hodinách výuky větší prostor pro výklad statistických metod a prezentaci jejich výstupů.

Dále bylo zjištěno, že méně pokročilí uživatelé informačních technologií spíše zastávají názor, že instruktážní prezentace ztrátlivnily jejich domácí přípravu. Naopak pokročilejší uživatelé počítače (experti) nemají s orientací v novém systémovém prostředí větší obtíže, a přestože existenci manuálů oceňují, není pro ně nezbytná. Tyto výsledky podporují cíle autorů IP, jejichž záměrem bylo směřovat pozornost především na méně zkušené uživatele informačních technologií, kteří mívají obecně větší problémy s orientací ve složitých programových systémech.

Literatura

- BEH, E.J. (2010). Elliptical confidence regions for simple correspondence analysis. *Journal of Statistical Planning and Inference* 140: 2582–2588. <http://dx.doi.org/10.1016/j.jspi.2010.03.018>
- HEBÁK, P., HUSTOPECKÝ, J., PECÁKOVÁ, I., PRŮŠA, M., ŘEZANKOVÁ, H., SVOBODOVÁ, A., VLACH, P. (2007). *Vícerozměrné statistické metody (3)*. Praha: Informatorium.
- MELOUN, M., MILITKÝ, J., HILL, M. (2005). *Počítačová analýza vícerozměrných dat v příkladech*. Praha: Academia.
- RAMOS, M., CARVALHO, H. (2010). Perceptions of quantitative methods in higher education: mapping student profiles. *Higher Education*, published online: 25. July 2010. Dostupné na WWW: <http://dx.doi.org/10.1007/s10734-010-9353-3>.
- RENCER, A. (2002). *Methods of Multivariate Analysis*. New York: Wiley.
- ŘEZANKOVÁ, H. (2007). *Analýza dat z dotazníkového šetření*. Praha: Profesional Publishing.